W1518

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-283123
(43)Date of publication of application : 23.10.1998

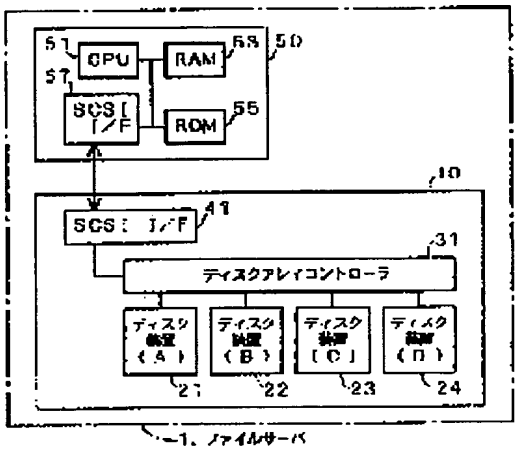
(51)Int.Cl. G06F 3/06
G06F 3/06
G06F 1/32
G06F 1/26

(21)Application number : 09-085364
(22)Date of filing : 03.04.1997

(71)Applicant : XING:KK
BROTHER IND LTD
(72)Inventor : MIYAKOSHI MITSUNARI
KUNO TAKUMA

(54) STORAGE DEVICE ARRAY

(57)Abstract:
PROBLEM TO BE SOLVED: To efficiently supply power and to suppress the degradation of a storage device and the consumption of the power by supplying the power to a standby device in the case of judging that an error is generated in one of minimum storage devices capable of generating the data of a read object.
SOLUTION: A disk array 10 is provided with disk devices 21-24 as the four storage devices to be parallelly operated and a disk array controller 31 capable of individually controlling the respective disk devices 21-24, etc. Then, in the disk array 10, control for supplying the power to the three disk devices 21-23 equivalent to the minimum storage devices capable of generating the data of the read object and not supplying the power to the disk device 24 equivalent to the standby device is normally performed. Then, in the case of a condition that the error is generated for one of the disk devices 21-23, the control is performed so as to supply the power to the disk device 24 for the first time.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision

of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号

特開平10－283123

(43)公開日　平成10年(1998)10月23日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ０６Ｆ	3/06	３０５	Ｇ０６Ｆ	3/06	３０５Ｃ
		５４０			５４０
	1/32			1/00	３３２Ｂ
	1/26				３３４Ｇ

審査請求　未請求　請求項の数４　ＯＬ　（全９頁）

(21)出願番号　特願平9－85364

(22)出願日　平成９年(1997)４月３日

(71)出願人　396004833
株式会社エクシング
名古屋市瑞穂区塩入町18番１号

(71)出願人　000005267
ブラザー工業株式会社
愛知県名古屋市瑞穂区苗代町15番１号

(72)発明者　宮腰　光成
愛知県名古屋市中区錦３丁目10番33号　株式会社エクシング内

(72)発明者　久野　琢磨
愛知県名古屋市中区錦３丁目10番33号　株式会社エクシング内

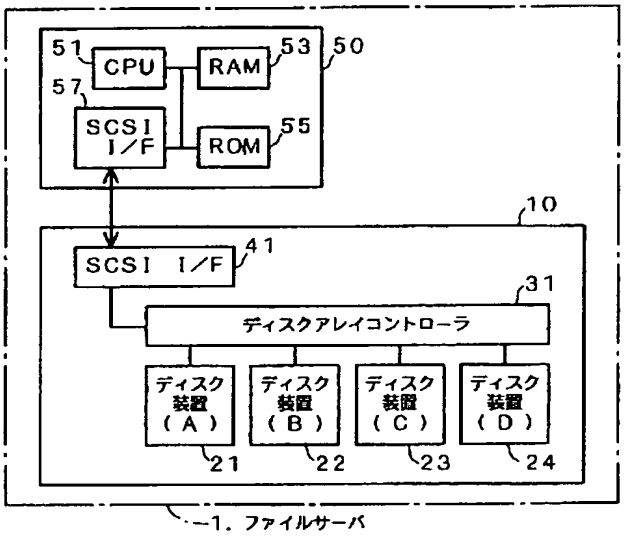
(74)代理人　弁理士　足立　勉

(54)【発明の名称】　記憶装置アレイ

(57)【要約】

【課題】　例えば複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能であり、残りは予備的な記憶装置として捉えることのできるＲＡＩＤにおいて、効率的な電源供給を行って、ディスク装置等の記憶装置の劣化及びその記憶装置による電力の消費を抑える。

【解決手段】　（Ａ）～（Ｃ）のディスク装置２１～２３の３台からデータが読み出せれば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的なディスク装置（Ｄ）２４については電源供給せず、エラーが発生して予備的なディスク装置（Ｄ）２４が必要となって初めて電源供給する。この電源供給制御はスイッチング制御部２５０を介したＣＰＵ２０１からのポート制御により、各電源ライン２５１，２５２，２５３，２５４を個別にＯＮ／ＯＦＦして実行する。



【特許請求の範囲】
【請求項１】　アレイコントローラによって複数台の記憶装置を並行してアクセス可能であり、当該複数台の記憶装置にデータを分散させて記憶し、当該複数台の記憶装置のうちの少なくとも１台の記憶装置には冗長データを記憶することによって、メインコンピュータからの指示によりデータ読出処理を行う際には、前記複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読出対象のデータを生成可能な記憶装置アレイにおいて、
前記並行してアクセス可能な記憶装置について個別にステータスを取得可能なステータス取得手段と、
前記読出対象のデータを生成可能な最低限の記憶装置についてのみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御する電源供給制御手段とを備え、
さらに、当該電源供給制御手段は、前記ステータス取得手段によって取得したステータスに基づいて、前記読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、前記予備装置への電源供給を行なうよう構成されており、
前記アレイコントローラは、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うよう構成されていることを特徴とする記憶装置アレイ。
【請求項２】　前記電源供給制御手段が、前記取得したステータスに基づき前記読出対象のデータを生成可能な最低限の記憶装置のいずれかでエラーが発生したと判断した場合に前記予備装置への電源供給を行ない、前記アレイコントローラが、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱う一連の電源供給制御処理を所定時間毎に実行するよう構成されていることを特徴とする請求項１に記載の記憶装置アレイ。
【請求項３】　検査用データを記憶しており、当該検査用データを用いて前記記憶装置へのデータ書き込み及び読み出しを実行した場合の前記ステータス取得手段によって取得したステータスに基づき、前記エラー発生の判断を実行するよう構成されていることを特徴とする請求項２に記載の記憶装置アレイ。
【請求項４】　前記ステータス取得手段は、プロトコルコントローラであり、前記並行してアクセス可能な記憶装置については個別にプロトコルコントローラが設けられていることを特徴とする請求項１乃至３のいずれかに記載の記憶装置アレイ。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、アレイコントローラによって複数台の記憶装置を制御し、メインコンピュータからのアクセスに対しては１台の記憶装置に見せかけて応答するよう構成された記憶装置アレイに関する。
【０００２】
【従来の技術】従来、高速アクセスを可能とし、かつ高信頼性を保証する記憶装置としてディスクアレイが提案されてきた。ディスクアレイは、例えばハードディスク装置等の小型のディスク装置を複数備えることで、大型ディスク装置に対抗して高速アクセスを可能にする方式として体系化されている。この体系化された方式はＲＡＩＤ（Redundant Arrays of Inexpensive Disks ）として、David A Pattersonら３人によって発表され（U.C.Berkeley Report NO.UCB/CSD 87/39参照）、現在ではレベル１からレベル５までが基本的な構成として考えられている。ここでは、本発明に関係する代表的なものとして、レベル３のＲＡＩＤについて説明する。
【０００３】レベル３のＲＡＩＤでは、１台のパリティ用ディスク装置と２台以上のデータ用ディスク装置からなる複数台のディスク装置を並行にアクセスできるようになっている。そして、メインコンピュータからの要求によってデータを記憶するときには、そのデータを１バイト単位に分割し、データ用ディスク装置のそれぞれに分散して記憶すると共に、それらの分割データの排他的論理和であるパリティデータを計算してパリティ用ディスク装置に記憶する。つまり、データは、並行してアクセスできる複数台のディスク装置のそれぞれに分散して記憶されることになる。
【０００４】このように、本来１台のディスク装置に書き込まれるはずのデータが複数台のディスク装置に分散して書き込まれるため、トータルで考えた場合、１台のディスク装置に対する読み書きの頻度や読み出すデータ量が低減されるため、全体として見れば処理速度の向上が図られるのである。また、複数台のディスク装置のうちの１台をパリティ用ディスク装置とするため、１台のデータ用ディスク装置からのデータが読み出せない場合であっても、残りのデータ用ディスク装置からのデータとパリティ用ディスク装置からのパリティデータに基づいて、その１台のデータ用ディスク装置から読み出されるはずのデータを生成することができ、信頼性の向上が図られるのである。
【０００５】そして、このディスクアレイ装置においては、データ読み出しのパフォーマンスの低下を防止するために、備えている複数台のディスク装置を全て稼働させて読み出しがいつでも開始できるようにされていた。ディスク装置の稼働中は電源が供給され続けている。
【０００６】
【発明が解決しようとする課題】しかしながら、このように全てのディスク装置を稼働させていても、常に全てのディスク装置からデータ読み出さなければメインコンピュータから要求されたデータを生成できないわけではない。例えば、上述したレエベル３のＲＡＩＤの場合でいえば全てのデータ用ディスク装置から正常にデータを

読みせればパリティデータがなくてもデータ生成ができるため、パリティ用ディスク装置からのデータ読み出しはしなくても構わなくなるのである。つまり、この場合のパリティ用ディスク装置は、予備的なディスク装置として捉えることができる。

【０００７】したがって、結果的には最終的なデータ生成には不要なディスク装置に対しても電源供給を行っていることとなり、ディスク装置の劣化が激しくなったり、また、無駄な電力が消費されるという問題が生じていた。本発明は、例えば複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能であり、残りは予備的な記憶装置として捉えることのできるＲＡＩＤにおいて、効率的な電源供給を行って、ディスク装置等の記憶装置の劣化及びその記憶装置による電力の消費を抑えることを目的とする。

【０００８】

【課題を解決するための手段及び発明の効果】上記目的を達成するためになされた請求項１に記載の記憶装置アレイは、アレイコントローラによって複数台の記憶装置を並行してアクセス可能であり、当該複数台の記憶装置にデータを分散させて記憶し、当該複数台の記憶装置のうちの少なくとも１台の記憶装置には冗長データを記憶することによって、メインコンピュータからの指示によりデータ読出処理を行う際には、前記複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読出対象のデータを生成可能な記憶装置アレイにおいて、前記並行してアクセス可能な記憶装置について個別にステータスを取得可能なステータス取得手段と、前記読出対象のデータを生成可能な最低限の記憶装置についてのみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御する電源供給制御手段とを備え、さらに、当該電源供給制御手段は、前記ステータス取得手段によって取得したステータスに基づいて、前記読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、前記予備装置への電源供給を行なうよう構成されており、前記アレイコントローラは、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うよう構成されていることを特徴とする。

【０００９】本発明の記憶装置アレイでは、アレイコントローラによって複数台の記憶装置を並行してアクセス可能である。なお、ｎ（ｎは２以上の自然数。以下、同じ。）台の記憶装置が並行してアクセス可能である場合、並行してアクセス可能なｎ系統に１台ずつの記憶装置が属し、ｎ台の記憶装置が並行してアクセス可能となっていることも考えられるし、並行してアクセス可能なｎ系統に２台以上の記憶装置が属し、その各系統から１台ずつを選択したｎ台の記憶装置が並行してアクセス可能となっていることも考えられる。

【００１０】アレイコントローラによって、データは複数台の記憶装置のそれぞれに分散して記憶され、その複数台の記憶装置のうちの少なくとも１台の記憶装置にはパリティデータ等の冗長データが記憶される。冗長データを利用することによって、メインコンピュータからの指示によるデータ読出処理では、複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読み出し対象のデータを生成することが可能である。例えばパリティデータが冗長データとして用いられる場合、所定数は（ｎ－１）となり、（ｎ－１）台の記憶装置からのデータに基づき読出対象のデータを生成可能である。

【００１１】そして、本記憶装置アレイでは、電源供給制御手段が、読出対象のデータを生成可能な最低限の記憶装置についてのみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御するのであるが、並行してアクセス可能な記憶装置についてステータス取得手段が個別にステータスを取得可能であり、電源供給制御手段は、その取得したステータスに基づいて、読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、予備装置への電源供給を行なう。そして、アレイコントローラは、エラーが発生した記憶装置の代わりに電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱う。

【００１２】このように、複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的な記憶装置については電源供給せず、エラーが発生して予備的な記憶装置が必要となって初めて電源供給することとなるので、効率的な電源供給を行える。したがって、データ読み出しに用いない場合の予備的な記憶装置の劣化及びその記憶装置による電力の消費を抑えることができる。

【００１３】なお、上述したように、本発明の記憶装置アレイでは、電源供給制御手段が、取得したステータスに基づき読出対象のデータを生成可能な最低限の記憶装置のいずれかでエラーが発生したと判断した場合に予備装置への電源供給を行ない、アレイコントローラが、エラーが発生した記憶装置の代わりに電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うという一連の「電源供給制御処理」を実行するのであるが、この処理は、例えば実際にメインコンピュータからの指示によりデータ読出処理を行う際に実行してもよいが、メインコンピュータからの指示によらず、記憶装置アレイ自身が所定時間毎に実行してもよい。

【００１４】つまり、メインコンピュータからの指示によりデータ読出処理を行う際に実行した場合には、エラーが生じて初めて予備装置への電源供給を開始し、その予備装置からデータを読み出して、読出対象のデータを生成することとなるため、レスポンスはどうしても低下

してしまう。したがって、そのようなレスポンスの低下の考慮が不要な時期に同様の制御処理を実行しておけばよい。そうすれば、前もってエラーが生じた記憶装置の代わりに予備装置をデータ読み出しの対象として扱う準備が完了するため、実際にメインコンピュータからの指示によりデータ読出処理を行う際のレスポンス低下を防止することができる。

【００１５】なお、このように、メインコンピュータからの指示によらず、記憶装置アレイ自身が所定時間毎に電源供給制御処理を実行する場合には、例えば、検査用データを記憶しており、当該検査用データを用いて記憶装置へのデータ書き込み及び読み出しを実行した場合のステータス取得手段によって取得したステータスに基づき、エラー発生の判断を実行するよう構成することが考えられる。

【００１６】また、ステータス取得手段は、例えばプロトコルコントローラを用い、並行してアクセス可能な記憶装置については個別にプロトコルコントローラを設けるように構成することが考えられる。例えば一般的なＲＡＩＤにおいては、記憶装置としてハードディスク装置が用いられ、アレイコントローラとハードディスク装置とをＳＣＳＩバスで接続しているため、ＳＣＳＩプロトコルコントローラ（ＳＰＣ）が採用されることが多い。

【００１７】

【発明の実施の形態】以下、本発明を具体化した一実施形態を図面を参照して説明する。図１は、本発明の記憶装置アレイをファイルサーバ１に適用した概略構成を示すブロック図である。

【００１８】図１に示すように、ファイルサーバ１は、「記憶装置アレイ」としてのディスクアレイ１０と、メインコンピュータ５０とから構成されている。メインコンピュータ５０は、制御手段としてのＣＰＵ５１と、「記憶手段」としてのＲＡＭ５３と、プログラム記憶手段としてのＲＯＭ５５と、ディスクアレイ１０と接続するためのＳＣＳＩインターフェース５７とを備えている。

【００１９】一方、ディスクアレイ１０は、並列に動作する４台の「記憶装置」としてのディスク装置２１，２２，２３，２４と、それら各ディスク装置２１～２４を個別に制御可能な「アレイコントローラ」としてのディスクアレイコントローラ３１と、メインコンピュータ５０と接続するためのＳＣＳＩインターフェース４１とを備えている。このように、並列に動作する４台のディスク装置２１～２４を備えるディスクアレイ１０を系統数４のディスクアレイ１０という。

【００２０】なお、以下の説明では、４台のディスク装置２１～２４を区別するために、ディスク装置（Ａ）２１、ディスク装置（Ｂ）２２、ディスク装置（Ｃ）２３及びディスク装置（Ｄ）２４と記載することにする。また、ディスク装置２１～２４は、いわゆる物理的なハードディスクドライブとそれを制御するコントロールボードとが一体化されたものである。

【００２１】続いて、ディスクアレイコントローラ３１の内部構成について図２を参照して説明する。なお、図２は、４台のディスク装置２１～２４に対する制御関連の構成だけを示したものであり、それ以外の構成は省略してある。ディスクアレイコントローラ３１は、本コントローラ全体の制御を司る制御手段としてのＣＰＵ２０１と、各種処理データを格納しておくＲＡＭ２００と、ＣＰＵ２０１が実行する動作プログラムなどを格納しておくＲＯＭ２０２と、４本のＳＣＳＩラインを構成するＳＣＳＩバス（Ａ）１０１、ＳＣＳＩバス（Ｂ）１０２、ＳＣＳＩバス（Ｃ）１０３及びＳＣＳＩバス（Ｄ）１０４に対してのリード及びライトデータをそれぞれ格納しておくための４つのバッファであるバッファ（Ａ）２１１、バッファ（Ｂ）２２１、バッファ（Ｃ）２３１及びバッファ（Ｄ）２４１と、各ＳＣＳＩラインとの間でのコマンド、データ及びステータスの送受信を行う４つのＳＣＳＩプロトコルコントローラ（ＳＰＣ）であるＳＰＣ（Ａ）２１３、ＳＰＣ（Ｂ）２２３、ＳＰＣ（Ｃ）２３３及びＳＰＣ（Ｄ）２４３とを備えている。

【００２２】さらに、ディスクアレイコントローラ３１は、各ＳＣＳＩラインからのデータリード及び各ＳＣＳＩラインへのデータライト時にデータをＤＭＡ（ダイレクト・メモリ・アクレス）転送するためのＤＭＡコントローラを各ライン毎に備えている。すなわち、ＤＭＡＣ（Ａ）２１２、ＤＭＡＣ（Ｂ）２２２、ＤＭＡＣ（Ｃ）２３１及びＤＭＡＣ（Ｄ）２４１である。

【００２３】また、各ＳＣＳＩラインには対応するディスク装置が接続されている。すなわち、ＳＣＳＩバス（Ａ）１０１にはディスク装置（Ａ）２１、ＳＣＳＩバス（Ｂ）１０２にはディスク装置（Ｂ）２２、ＳＣＳＩバス（Ｃ）１０３にはディスク装置（Ｃ）２３、そしてＳＣＳＩバス（Ｄ）１０４にはディスク装置（Ｄ）２４が接続されている。

【００２４】また、４本のＳＣＳＩラインであるＳＣＳＩバス（Ａ）１０１、ＳＣＳＩバス（Ｂ）１０２、ＳＣＳＩバス（Ｃ）１０３及びＳＣＳＩバス（Ｄ）１０４に対応する電源ライン２５１，２５２，２５３，２５４は、ＣＰＵ２０１のポートに割り振られており、スイッチング制御部２５０を介してＣＰＵ２０１からのポート制御により、各電源ライン２５１，２５２，２５３，２５４を個別にＯＮ／ＯＦＦすることができるようにされている。

【００２５】ここで、最初に本実施形態のファイルサーバ１の動作の前提となる基本的な機能を説明する。本実施形態のファイルサーバ１においては、メインコンピュータ５０から転送されたデータをディスクアレイコントローラ３１が１バイト単位に分割し、この場合は、ディスク装置（Ａ）２１、ディスク装置（Ｂ）２２、ディス

ク装置（Ｃ）２３の３台の同一セクタ上に順次書き込むと共に、それらのデータに対応するパリティを生成してディスク装置（Ｄ）２４に書き込むよう構成されている。

【００２６】従って、読出対象となるデータは、（Ａ）～（Ｃ）の３台のディスク装置２１～２３からのデータを結合することにより生成できる。このとき、（Ａ）～（Ｃ）の３台のディスク装置２１～２３のうちの１台にリードエラーが発生しても、残りの２台のディスク装置とディスク装置（Ｄ）２４からのパリティデータに基づいて、リードエラーとなったディスク装置から読み出されるはずのデータを生成することが可能であり、読出対象となるデータを生成することができる。ファイルサーバ１のこのような基本的な機能を考えると、（Ａ）～（Ｃ）のディスク装置２１～２３の３台が「読出対象のデータを生成可能な最低限の記憶装置」に相当し、ディスク装置（Ｄ）２４が「予備装置」に相当することとなり、本実施形態のディスクアレイ１０はレベル３のＲＡＩＤに相当する。

【００２７】したがって、本実施形態のディスクアレイ１０では、読出対象のデータを生成可能な最低限の記憶装置に相当する（Ａ）～（Ｃ）のディスク装置２１～２３の３台については電源供給し、「予備装置」に相当するディスク装置（Ｄ）２４ついては電源供給を行わないよう制御を通常は行っている。そして、（Ａ）～（Ｃ）のディスク装置２１～２３の内のいずれか１台についてエラーが生じる状況となった場合に初めてディスク装置（Ｄ）２４について電源供給を行うように制御するのである。これらの電源供給の実際の制御はＣＰＵ２０１が行なうのであるが、上述したように、スイッチング制御部２５０を介したＣＰＵ２０１からのポート制御により、各電源ライン２５１，２５２，２５３，２５４を個別にＯＮ／ＯＦＦしてこの電源供給制御を実行することとなる。

【００２８】図３には、この電源供給、停止にかかる判断及び指示内容を示すフローチャートである。本処理は、メインコンピュータ５０からのデータ読み出し要求などがない状態において所定時間毎に実行される。この処理ルーチンは、現在使用しているＳＣＳＩラインに対応するＳＰＣからのステータスをチェックし、その結果に応じて電源供給の切替などを制御するルーチンである。ここでは、上述したように（Ａ）～（Ｃ）のディスク装置２１～２３の３台については電源供給し、ディスク装置（Ｄ）２４については電源供給を行わないような制御を基本的に実行していると前提とする。したがって、ＳＣＳＩバス（Ａ）１０１、ＳＣＳＩバス（Ｂ）１０２、ＳＣＳＩバス（Ｃ）１０３の３本のバスにはそれぞれ対応する電源ライン２５１，２５２，２５３がＯＮに制御されていて電源供給がされており、ＳＣＳＩバス（Ｄ）１０４に対応する電源ライン２５４はＯＦＦに制御されていて電源供給がされていない状況である。したがって、この場合には、ＳＰＣ（Ａ）２１３、ＳＰＣ（Ｂ）２２３及びＳＰＣ（Ｃ）２３３からのステータスをチェックすることとなる。

【００２９】最初のステップＳ１では、ＣＰＵ２０１はＳＰＣ（Ａ）２１３から入力したステータスをチェックしてエラー発生がどうかを判断する。なお、この場合には、例えばＲＡＭ２００に所定の検査用データを記憶しており、その検査用データを用いてディスク装置（Ａ）２１へのデータ書き込み及び読み出しを実行し、その際のＳＰＣ（Ａ）２１３から入力したステータスに基づいてエラー発生の判断を実行する。なお、この処理に際しては、ＲＡＭ２００内の検査用データはバッファ（Ａ）２１１に送られ、ＤＭＡＣ（Ａ）２１２によってＳＰＣ（Ａ）２１３に転送されることとなる。

【００３０】そして、エラー発生の場合には（Ｓ１：ＹＥＳ）、Ｓ４へ移行して、対応するＳＣＳＩバス（Ａ）１０１への電源供給をＯＦＦするようにスイッチング制御部２５０に指示する。続くＳ７では、予備装置であるディスク装置（Ｄ）２４が接続されているＳＣＳＩバス（Ｄ）１０４への電源供給をＯＮするようにスイッチング制御部２５０に指示する。そして、Ｓ８では、上記Ｓ７で電源供給を開始したＳＣＳＩバス（Ｄ）に接続されているディスク装置（Ｄ）２４を、エラーが発生したＳＣＳＩバス（Ａ）１０１に接続されていたディスク装置（Ａ）２１の代わりに、読出対象データの生成用のディスク装置として扱うように設定してから、本処理ルーチンを終了する。

【００３１】一方、ＳＰＣ（Ａ）２１３から入力したステータスに対するチェックではエラー発生がなかった場合には（Ｓ１：ＮＯ）、Ｓ２へ移行して、今度はＳＰＣ（Ｂ）２２３から入力したステータスに対するチェックを行い、上述したＳＰＣ（Ａ）２１３の場合と同様の処理を行なう。すなわち、エラー発生の場合には（Ｓ２：ＹＥＳ）、Ｓ５へ移行して、対応するＳＣＳＩバス（Ｂ）１０２への電源供給をＯＦＦしてからＳ７へ移行する。Ｓ７以降の処理は上述したので詳しい説明は省略するが、Ｓ８では、この場合、ディスク装置（Ｄ）２４を、エラーが発生したＳＣＳＩバス（Ｂ）１０２に接続されていたディスク装置（Ｂ）２２の代わりに、読出対象データの生成用のディスク装置として扱うこととなる。

【００３２】また、ＳＰＣ（Ｂ）２２３から入力したステータスに対するチェックではエラー発生がなかった場合には（Ｓ２：ＮＯ）、Ｓ３へ移行して、今度はＳＰＣ（Ｃ）２３３から入力したステータスに対するチェックを行い、エラー発生の場合には（Ｓ３：ＹＥＳ）、Ｓ６へ移行して、対応するＳＣＳＩバス（Ｃ）１０３への電源供給をＯＦＦしてからＳ７、Ｓ８の処理を実行する。この場合、Ｓ８では、ディスク装置（Ｄ）２４を、エラ

ーが発生したＳＣＳＩバス（Ｃ）１０３に接続されていたディスク装置（Ｃ）２３の代わりに、読出対象データの生成用のディスク装置として扱うこととなる。

【００３３】もちろん、ＳＰＣ（Ｃ）２３３から入力したステータスに対するチェックでもエラー発生がなかった場合には（Ｓ３：ＮＯ）、そのまま本処理ルーチンを終了する。なお、ＣＰＵ２０１が「電源供給制御手段」に相当し、上述の図３のフローチャートにて示した処理が電源供給制御手段としての処理に相当する。

【００３４】このように、本実施形態のディスクアレイ１０においては、（Ａ）～（Ｃ）のディスク装置２１～２３の３台からデータが読み出せれば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的なディスク装置（Ｄ）２４については電源供給せず、エラーが発生して予備的なディスク装置（Ｄ）２４が必要となって初めて電源供給することとなるので、効率的な電源供給を行える。したがって、データ読み出しに用いない場合の予備的なディスク装置（Ｄ）２４の劣化及びそのディスク装置（Ｄ）２４及び接続されているＳＣＳＩバス（Ｄ）１０４による電力消費を抑えることができる。

【００３５】以上、本発明はこのような実施形態に何等限定されるものではなく、本発明の主旨を逸脱しない範囲において種々なる形態で実施し得る。例えば、上記実施形態のディスクアレイ１０においては、メインコンピュータ５０からの指示によらず、ディスクアレイ１０自身が所定時間毎に電源供給制御を実行するようにしたが、実際にメインコンピュータ５０からの指示によりデータ読出処理を行う際に実行してもよい。つまり、上述した実施形態では、ＲＡＭ２００内の検査用データを用いてディスク装置へのデータ書き込み及び読み出しを実行し、その際のＳＰＣから入力したステータスに基づいてエラー発生の判断を実行したが、メインコンピュータ５０から指示されたデータを読み出し、その際のＳＰＣから入力したステータスに基づいてエラー発生の判断を実行すればよい。

【００３６】但し、メインコンピュータ５０からの指示によりデータ読出処理を行う際に実行した場合には、エラーが生じて初めて予備装置（上記実施形態ではディスク装置（Ｄ）２４）の接続されているＳＣＳＩバス（Ｄ）１０４への電源供給を開始し、そのディスク装置（Ｄ）２４からデータを読み出して、読出対象のデータを生成することとなるため、レスポンスはどうしても低下してしまう。したがって、上記実施形態のようにすれば、レスポンスの低下の考慮が不要な時期に同様の制御処理を実行することができる。つまり、前もってエラーが生じたディスク装置の代わりに予備装置であるディスク装置（Ｄ）２４をデータ読み出しの対象として扱う準備が完了するため、実際にメインコンピュータ５０からの指示によりデータ読出処理を行う際のレスポンス低下を防止することができるのである。

【００３７】なお、上記実施形態では、４本のＳＣＳＩバス１０１～１０４に対してそれぞれ１台ずつのディスク装置２１～２４が接続される構成であったが、各ＳＣＳＩバス１０１～１０４に対してそれぞれ２台以上のディスク装置を接続するような構成であってもよい。

【００３８】また、上記実施形態では、レベル３のＲＡＩＤについて説明したが、その他のレベルのＲＡＩＤであっても、「読出対象のデータを生成可能な最低限の記憶装置」以外に「予備装置」として捉えることができる装置を備える構成の場合には同様に適用できる。

【００３９】ここでは、上述したレベル３も含めてディスクアレイのその他のレベルについて簡単に説明しておく。以下の説明では、ディスクを何台並列に並べるかという数を、パラレル数と呼び、ｐという変数で表すこととする。但し、パラレル数ｐには、冗長データを格納するためのディスクは含めない。

【００４０】なお、図４（Ａ）にはレベル０のＲＡＩＤ、（Ｂ）にはレベル１のＲＡＩＤの原理説明図を示し、図５（Ａ）にはレベル２のＲＡＩＤ、（Ｂ）にはレベル３のＲＡＩＤの原理説明図を示している。また、図６（Ａ）にはレベル４のＲＡＩＤ、（Ｂ）にはレベル５のＲＡＩＤの原理説明図を示している。

【００４１】ＲＡＩＤレベル０は、単にディスクをパラレル動作させ、データを分散記憶させるものである。信頼性向上の効果はなく、高速化の効果しかない。分散の単位は、ビット単位でもバイト単位でもセクタ単位でも良く、特に限定されない。ＲＡＩＤレベル０は正確にはＲＡＩＤではないが、対比のためによく例に挙げられる。

【００４２】ＲＡＩＤレベル１は、ミラーリングとも呼ばれ、２つのディスクに同一のデータを書き込み、読み出すときはどちらか一方のディスクから読み出す。片方のディスクが壊れても、データは失われない。ＲＡＩＤレベル１は高速化の効果はないが信頼性が増大する。

【００４３】ＲＡＩＤレベル２は、ハミングコードなどの冗長符号（誤り訂正符号）を用いるもので、レベル１が通常のディスクと比べて２倍のディスクを必要とするのに対し、レベル２は２倍までは要らない。データはビット単位あるいはバイト単位で分散させ、冗長符号と合わせて記録する。冗長符号の選び方で色々な方法が考えられるが特に冗長符号として単純なパリティを用いたものが次のレベル３である。他の冗長符号を用いたものはあまり実用価値がないので、レベル２が使用されることは希である。

【００４４】ＲＡＩＤレベル３は、バイト単位でデータを分散させ、パリティを付加してディスクに格納するもので、高速化、信頼性向上の両方の効果がある。反面、ディスクのセクタサイズ×ｐ個のデータが集まらないと読み書きができないので、ディスクを単体で使う場合に

比べて、アクセス単位が大きくなる。データを小さな単位でアクセスするような用途には不向きである。

【００４５】ＲＡＩＤレベル４は、この点を改善したもので、ディスクのセクタ単位でデータを分散させるものである。データはセクタサイズ単位に分割して、各ディスクに順に格納するのだが、ｐ個のセクタごとにパリティデータを計算して冗長データディスクに格納する。ＲＡＩＤレベル３ではセクタ単位×ｐのデータが最小アクセス単位だったが、ＲＡＩＤレベル４ではセクタ単位で読み書きが可能である。しかしセクタを書き換えるときは元のパリティデータと元のディスクデータを読み出して、新パリティを計算し直し、これをパリティとして書き戻す必要があるため、通常より余分な動作が必要となる。またパリティを格納しているディスクにアクセスが集中するため、ここがボトルネックになるという欠点もある。

【００４６】ＲＡＩＤレベル５はこの点を改善したもので、パリティを格納するディスクを、ブロックごとに回転させることで、特定のディスクにアクセスが集中するのを防止している。さらには、ＲＡＩＤ以外のものであっても、「読出対象のデータを生成可能な最低限の記憶装置」以外に「予備装置」として捉えることができる装置を備える構成の場合には同様に適用できる。

【図面の簡単な説明】

【図１】　本発明の記憶装置アレイをファイルサーバに適用した概略構成を示すブロック図である。

【図２】　実施形態のディスクアレイコントローラの内部構成を示すブロック図である。

【図３】　実施形態のディスクアレイコントローラが実行する電源供給制御処理を示すフローチャートである。

【図４】　ＲＡＩＤレベル０及びＲＡＩＤレベル１の原理説明図である。

【図５】　ＲＡＩＤレベル２及びＲＡＩＤレベル３の原理説明図である。

【図６】　ＲＡＩＤレベル４及びＲＡＩＤレベル５の原理説明図である。

【符号の説明】

１…ファイルサーバ　　１０…ディスクアレイ

２１～２４…ディスク装置　　３１…ディスクアレイコントローラ

４１…ＳＣＳＩインターフェース　　５０…メインコンピュータ

５１…ＣＰＵ　　５３…ＲＡＭ

５５…ＲＯＭ　　５７…ＳＣＳＩインターフェース

１０１～１０４…ＳＣＳＩバス　　２００…ＲＯＭ

２０１…ＣＰＵ　　２０２…ＲＡＭ

２１１，２２１，２３１，２４１…バッファ

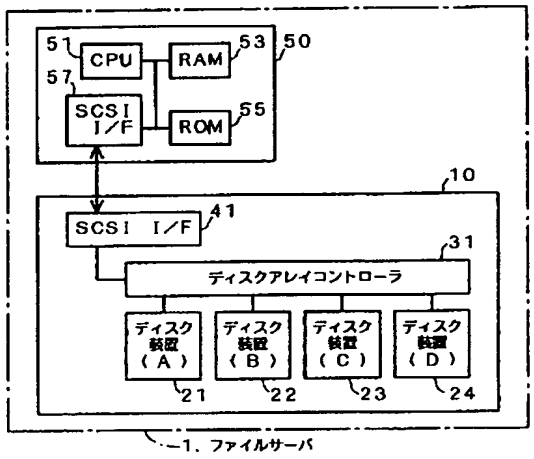
２１２，２２２，２３２，２４２…ＤＭＡコントローラ

２１３，２２３，２３３，２４３…ＳＣＳＩプロトコルコントローラ（ＳＰＣ）

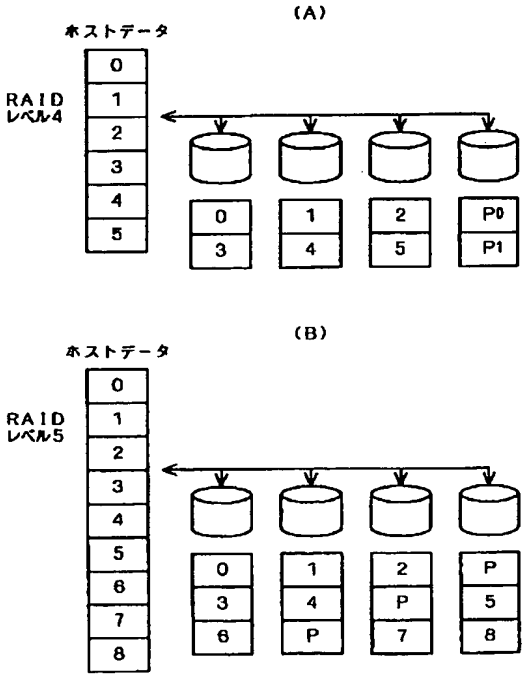
２５０…スイッチング制御部

２５１～２５４…電源ライン

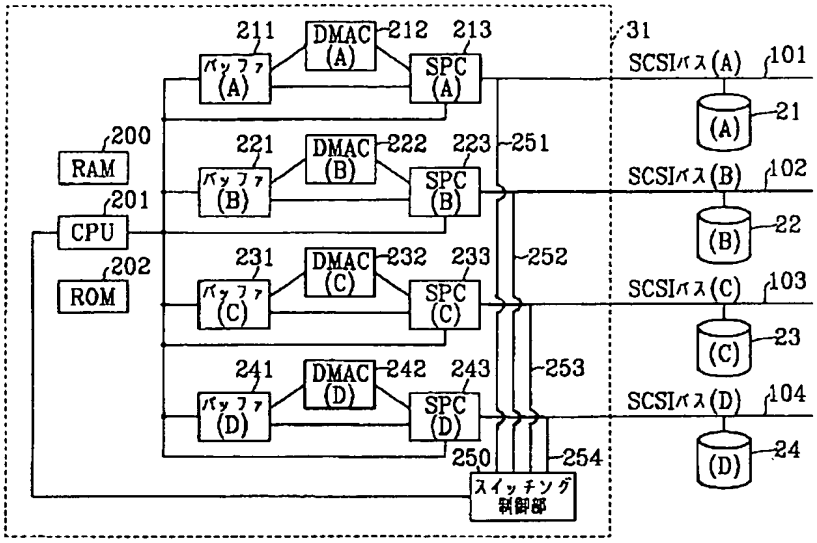
【図１】



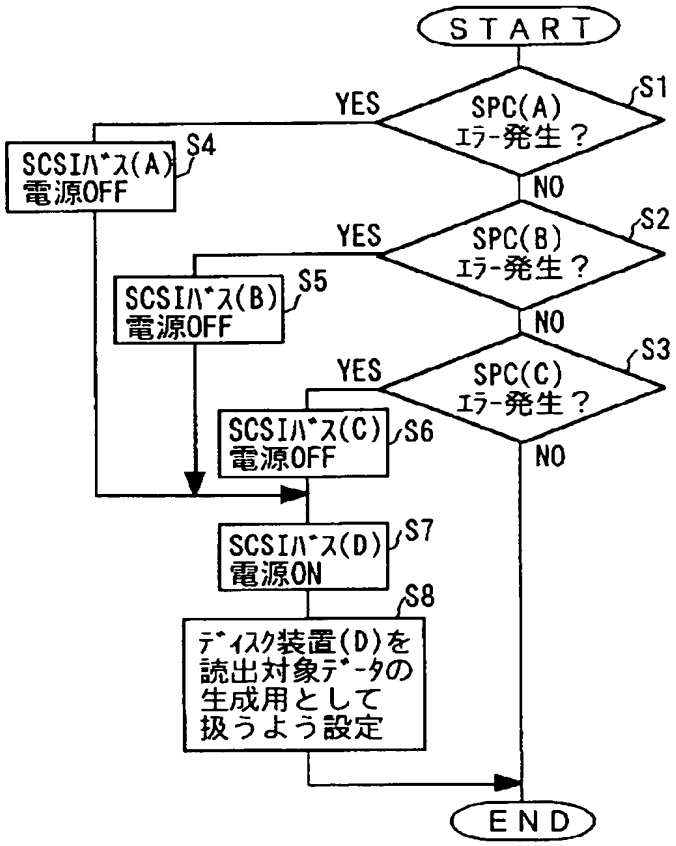
【図６】



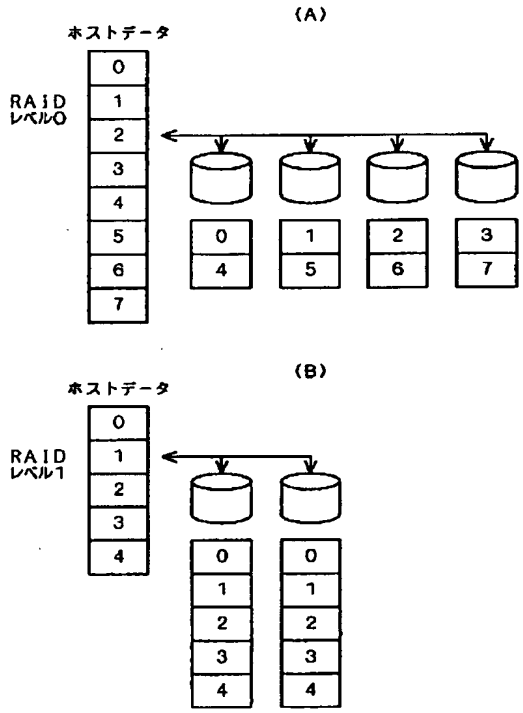
【図２】


31
211
DMAC (A)
212
213
バッファ (A)
SPC (A)
SCSIバス(A)
101
(A)
21
200
RAM
221
DMAC (B)
222
223
251
バッファ (B)
SPC (B)
SCSIバス(B)
102
201
CPU
(B)
22
231
DMAC (C)
232
233
252
202
ROM
バッファ (C)
SPC (C)
SCSIバス(C)
103
(C)
23
241
DMAC (D)
242
243
253
バッファ (D)
SPC (D)
SCSIバス(D)
104
250
254
(D)
24
スイッチング制御部

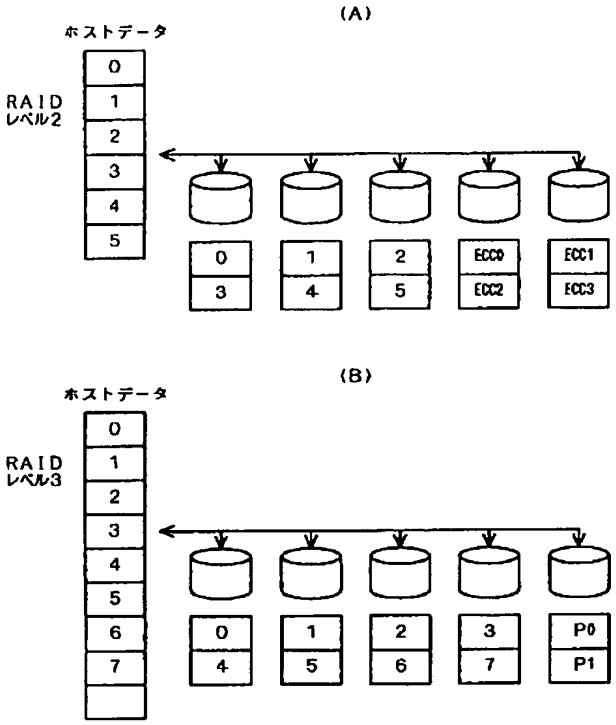
【図３】


START
SPC(A) エラー発生？
S1
YES
NO
SCSIバス(A) 電源OFF
S4
SPC(B) エラー発生？
S2
YES
NO
SCSIバス(B) 電源OFF
S5
SPC(C) エラー発生？
S3
YES
NO
SCSIバス(C) 電源OFF
S6
SCSIバス(D) 電源ON
S7
S8
ディスク装置(D)を読出対象データの生成用として扱うよう設定
END

【図４】


(A)
ホストデータ
RAID レベル0
0
1
2
3
4
5
6
7
0
4
1
5
2
6
3
7
(B)
ホストデータ
RAID レベル1
0
1
2
3
4
0
1
2
3
4
0
1
2
3
4

【図５】


(A)
ホストデータ
RAID
レベル2
0
1
2
3
4
5
0
3
1
4
2
5
ECC0
ECC2
ECC1
ECC3
(B)
ホストデータ
RAID
レベル3
0
1
2
3
4
5
6
7
0
4
1
5
2
6
3
7
P0
P1